SUBSTITUTE SPECIFICATION – CLEAN VERSION

Application No.: 10/531,632

METHOD FOR PROCESSING 3-D AUDIO SCENE WITH EXTENDED SPATIALITY OF SOUND SOURCE

Description

Technical Field

The present invention relates to a method for processing a three-dimensional audio scene having sound source whose spatiality is extended; and, more particularly, to a method for processing a three-dimensional audio scene to extend the spatiality of sound source in a three-dimensional audio scene.

Background Art

Generally, a content providing server encodes contents in a predetermined encoding method and transmits the encoded contents to content consuming terminals that consume the contents. The content consuming terminals decode the contents in a predetermined decoding method and output the transmitted contents.

Accordingly, the content providing server includes an encoding unit for encoding the contents and a transmission unit for transmitting the encoded contents. On the other hand, the content consuming terminals includes a reception unit for receiving the transmitted encoded contents, a decoding unit for decoding the encoded contents, and an output unit for outputting the decoded contents to users.

Many encoding/decoding methods of audio/video signals are known so far. Among them, an encoding/decoding method based on Moving Picture Experts Group 4 (MPEG-4) is widely used these days. MPEG-4 is a technical standard for data compression and restoration technology defined by the MPEG to transmit moving pictures at a low transmission rate.

MPEG-4, which is ISO/IEC 14496-1, defines technology

for coding/decoding audio visual (AV) scene in terms of audio visual information and associated scene description information. The entity that composes and sends, or receives and presents such a coded representation of an audio visual scene is generically referred to as an "audio visual terminal" or just "terminal". This terminal may correspond to a stand-alone application or be part of an application system.

The MPEG-4 describes a system for communicating audio visual information, that is, the representation of physical or virtual objects that can be manifested audibly and/or visually. At the encoder side, audio visual information related to a physical scene is compressed, error protected if necessary and multiplexed in one or more coded binary streams that are transmitted. At the decoder side, these streams are demultiplexed, error corrected, decompressed, composited, and presented to the end user. This is revealed in "Coding Of Moving Pictures And Audio", ISO/IEC JTC1/SC29/WG11 N1483, Systems Working Draft Version 2.0, November 22, 1996).

According to MPEG-4, an object of an arbitrary shape can be encoded and the content consuming terminals consume a scene composed of a plurality of objects. Therefore, MPEG-4 defines Audio Binary Format for Scene (Audio BIFS) with a scene description language for designating a sound object expression method and the characteristics thereof.

Scene description means information that describes the spatio-temporal positioning of audio visual objects, and this is revealed in "Coding Of Moving Pictures And Audio," ISO/IEC JTC1/SC29/WG11 N1901, Text for CD 14496-1 Systems, November 21, 1997. MPEG-4, which is ISO/IEC 14496, addresses the coding of audio visual objects of various types: natural video and audio objects, and also synthetic music and sound effects. To reconstruct a multimedia scene at the terminal, it is hence not sufficient to transmit the

raw audio visual data to a receiving terminal. Additional information is needed in order to combine this audio visual data at the terminal and construct and present to the end user a meaningful multimedia scene. This information, called scene description, determines the placement of audio visual objects in space and time and is transmitted together with the objects. The scene description only describes the structure of the scene. The action of assembling these objects in the same representation space is called composition. The action of transforming these audio visual objects from a common representation space to a specific presentation device, i.e., speakers and a viewing window, is called rendering.

Examples of AV objects are conventional video, conventional audio, pre-mixed audio tracks which include speech, music, synthetic audio such as MIDI, and the like. AV objects can be hierarchical in a sense that they may be defined as composites of other AV objects, which are called sub-objects. AV objects that are composites of sub-objects are called compound AV objects. All other AV objects are called primitive AV objects. The top-most object in the hierarchy is called the "scene". An AV Scene is the topmost AV object in the hierarchy of compound AV objects, and this is revealed in "Coding Of Moving Pictures And Audio," ISO/IEC JTC1/SC29/WG11 N1483, Systems Working Draft Version 2.0, November 22, 1996. A set of AV objects is called an AV scene and an AV scene includes scene description information which defines spatiotemporal attributes of the AV objects.

Meanwhile, along with the development in video, users want to consume contents of more lifelike sounds and video quality. In the MPEG-4 Audio Binary Format for Scene (Audio BIFS), an AudioFX node and a DirectiveSound node are used to express spatiality of a three-dimensional audio scene.

A BIFS scene description is a compact binary format representing a pre-defined set of objects. The BIFS format contains information on the attributes of objects, which define their audio visual properties and the structure of the scene graph which contains these objects. The BIFS scene description data is itself conveyed to the receiver as an Elementary Stream.

The BIFS scene description includes a collection of nodes which describe the scene and its layout. An object in the scene is described by one or more nodes, which may be grouped together using a grouping node.

An object may be completely described within the BIFS information, or may also require streaming data from one or more AV decoders. In the latter case, the node points out an object descriptor or a URL descriptor which indicates which Elementary Stream(s) is (are) associated with the node.

Sound nodes are used for building audio scenes from audio sources coded with MPEG-4 coding tools. Sound may be included in either the 2D or 3D scene graphs. In a 3D scene, the sound may be spatially presented to apparently originate from a particular 3D direction, according to the positions of the object and the listener. The Sound node is used to attach sound to 3D and 2D scene graphs. As with visual objects, the audio objects represented by this node has a position in space and time, and are transformed by the spatial and grouping transforms of nodes hierarchically above them in the scene. The nodes below the Sound nodes, however, constitute an audio subtree. This subtree is used to describe a particular audio object through the mixing and processing of several audio streams. Rather than representing a hierarchy of spatiotemporal tranformations,

the nodes within the audio subtree represent a signal-flow graph that describes how to create the audio object from the sounds coded in the AudioSource streams. That is, each of audio subtree nodes, i.e., AudioSource, AudioMix, AudioSwitch, AudioFX, Sound, etc., accepts one or several channels of input sound, and describes how to turn these channels of input sound into one or more channels of output sound. The only sounds presented in the audio-visual scene are those sounds which are the output of audio nodes that are children of a Sound node, that is, the "highest" outputs in the audio subtree. Herein, the AudioSource defines sound input for a scene and the AudioMix mixes sound. The AudioSwitch switches sound in a scene, and the AudioFX combines audio objects configured for sound which defines particular sound. The Sound defines properties of sound. The AudioSource is used to add sound to a scene. Diverse audio coding tools are revealed in the ISO/IEC CD 14496-3:1997. The audio nodes present in an audio subtree do not each represent a sound to be presented in the scene. Rather, the audio subtree represents a signal-flow graph which computes a single (possibly multichannel) audio object based on a set of audio inputs in AudioSource nodes and parametric transformations. The only sounds which are presented to the listener are those which are the "output" of these audio subtrees, as connected to Sound node. This is revealed in "Coding Of Moving Pictures And Audio" 1997.11.21. ISO/IEC JTC1/SC29/WG11 N1901, Text for CD 14496-1 Systems.

In these nodes, modeling of sound source is usually depended on point-source. Point-source can be described and embodied in a three-dimensional sound space easily.

Actual point-sources, however, tend to have a dimension more than two, rather than to be a point of

literal meaning. More important thing here is that the shape of the sound source can be recognized by human beings, which is disclosed by J. Baluert, "Spatial Hearing," the MIT Press, Cambridge Mass, 1996.

For example, a sound of waves dashing against the coastline stretched in a straight line can be recognized as a linear sound source instead of a point sound source.

To improve the sense of the real of the three-dimensional audio scene by using the Audio BIFS, the size and shape of the sound source should be expressed. Otherwise, the sense of the real of a sound object in the three-dimensional audio scene would be damaged seriously.

That is, the spatiality of a sound source could be described to endow a three-dimensional audio scene with a sound source which is of more than one-dimensional.

Disclosure of Invention

It is, therefore, an object of the present invention to provide a method for processing a three-dimensional audio scene having a sound source whose spatiality is extended by adding sound source characteristics information having information on extending the spatiality of the sound source to three-dimensional audio scene description information.

The other objects and advantages of the present invention can be easily recognized by those of ordinary skill in the art from the drawings, detailed description and claims of the present specification.

In accordance with one aspect of the present invention, there is provided a method for processing a three-dimensional audio scene with a sound source whose spatiality is extended, including the steps of: a) generating 3D audio scene description information including sound source characteristics information of a sound object;

and b) coding the sound object and the 3D audio scene description information including the sound source characteristics information of the sound object, wherein the sound source characteristics information includes spatiality extension information of the sound source which is information on the size and shape of the sound source expressed in a three-dimensional space.

In accordance with another aspect of the present invention, there is provided a method for processing a three-dimensional audio scene with a sound source whose spatiality is extended, which includes the steps of: a) decoding a sound object and 3D audio scene description information including sound source characteristics information for the sound object; and b) outputting the sound object based on the three-dimensional audio scene description information, wherein the sound source characteristics information includes spatiality extension information which is information on the size and shape of the sound source expressed in a three-dimensional space.

In accordance with another aspect of the present invention, there is provided a three-dimensional audio scene data stream with a sound source whose spatiality is extended, which includes: a sound object; and three-dimensional audio scene description information including sound source characteristics information for the sound object data, wherein the sound source characteristics information includes spatiality extension information which is information on the size and shape of the sound source expressed in a three-dimensional space.

Brief Description of Drawings

The above and other objects and features of the present invention will become apparent from the following description of the preferred embodiments given in

conjunction with the accompanying drawings, in which:

Fig. 1 is a diagram illustrating various shapes of sound sources;

Fig. 2 is a diagram describing a method for expressing spatial sound source by grouping successive point sound sources;

Fig. 3 shows an example where spatiality extension information is added to a "DirectiveSound" node of Audio BIFS in accordance with the present invention;

Fig. 4 is a diagram illustrating how a sound source is extended in accordance with the present invention; and

Fig. 5 is a diagram depicting the distributions of point sound sources based on the shapes of various sound sources in accordance with the present invention.

Best Mode for Carrying Out the Invention

Other objects and aspects of the invention will become apparent from the following description of the embodiments with reference to the accompanying drawings, which is set forth hereinafter.

Following description exemplifies only the principles of the present invention. Even if they are not described or illustrated clearly in the present specification, one of ordinary skill in the art can embody the principles of the present invention and invent various apparatuses within the concept and scope of the present invention.

The use of the conditional terms and embodiments presented in the present specification are intended only to make the concept of the present invention understood, and they are not limited to the embodiments and conditions mentioned in the specification.

In addition, all the detailed description on the principles, viewpoints and embodiments and particular embodiments of the present invention should be understood

to include structural and functional equivalents to them. The equivalents include not only currently known equivalents but also those to be developed in future, that is, all devices invented to perform the same function, regardless of their structures.

For example, block diagrams of the present invention should be understood to show a conceptual viewpoint of an exemplary circuit that embodies the principles of the present invention. Similarly, all the flowcharts, state conversion diagrams, pseudo codes and the like can be expressed substantially in a computer-readable media, and whether or not a computer or a processor is described distinctively, they should be understood to express various processes operated by a computer or a processor.

Functions of various devices illustrated in the drawings including a functional block expressed as a processor or a similar concept can be provided not only by using hardware dedicated to the functions, but also by using hardware capable of running proper software for the functions. When a function is provided by a processor, the function may be provided by a single dedicated processor, single shared processor, or a plurality of individual processors, part of which can be shared.

The apparent use of a term, 'processor', 'control' or similar concept, should not be understood to exclusively refer to a piece of hardware capable of running software, but should be understood to include a digital signal processor (DSP), hardware, and ROM, RAM and non-volatile memory for storing software, implicatively. Other known and commonly used hardware may be included therein, too.

In the claims of the present specification, an element expressed as a means for performing a function described in the detailed description is intended to include all methods for performing the function including all formats of software, such as combinations of circuits for performing

the intended function, firmware/microcode and the like. To perform the intended function, the element is cooperated with a proper circuit for performing the software. The present invention defined by claims includes diverse means for performing particular functions, and the means are connected with each other in a method requested in the claims. Therefore, any means that can provide the function should be understood to be an equivalent to what is figured out from the present specification.

Other objects and aspects of the invention will become apparent from the following description of the embodiments with reference to the accompanying drawings, which is set forth hereinafter. The same reference numeral is given to the same element, although the element appears in different drawings. In addition, if further detailed description on the related prior arts is determined to blur the point of the present invention, the description is omitted. Hereafter, preferred embodiments of the present invention will be described in detail.

Fig. 1 is a diagram illustrating various shapes and sizes of sound sources. Referring to Fig. 1, a sound source can be a point, a line, a surface and space having a volume. Since sound source has an arbitrary shape and size, it is very complicated to describe the sound source. However, if the shape of the sound source to be modeled is controlled, the sound source can be described less complicatedly.

In the present invention, it is assumed that point sound sources are distributed uniformly in the dimension of a virtual sound source in order to model sound sources of various shapes and sizes. As a result, the sound sources of various shapes and sizes can be expressed as continuous arrays of point sound sources. Here, the location of each point sound source in a virtual object can be calculated using a vector location of a sound source which is defined

in a three-dimensional scene.

When a spatial sound source is modeled with a plurality of point sound sources, the spatial sound source should be described using a node defined in Audio BIFS. When the node defined in Audio BIFS, which will be referred to as an AudioFX node, is used, any effect can be included in the three-dimensional scene. Therefore, an effect corresponding to the spatial sound source can be programmed through the AudioFX node and inserted to the three-dimensional scene.

However, this requires very complicated Digital Signal Processing (DSP) algorithm and it is very troublesome to control the dimension of the spatial sound source.

Also, the point sound sources distributed in a limited dimension of an object are grouped using the Audio BIFS, and the spatial location and direction of the sound sources can be changed by changing the sound source group. First of all, the characteristics of the point sound sources are described using a plurality of "DirectiveSound" node. The locations of the point sound sources are calculated to be distributed on the surface of the object uniformly.

Subsequently, the point sound sources are located with a spatial distance that can eliminate spatial aliasing, which is disclosed by A. J. Berkhout, D. de Vries, and P. Vogel, "Acoustic control by wave field synthesis," J. Aoust. Soc. Am., Vol. 93, No. 5 on pages from 2764 to 2778, May, 1993. The spatial sound source can be vectorized by using a group node and grouping the point sound sources.

Fig. 2 is an illustrative diagram depicting a scene of Audio BIFS. In the drawing, a virtual successive linear sound source is modeled by using three point sound sources which are distributed uniformly along the axis of the linear sound source.

The locations of the point sound sources are

determined to be (x_0-dx, y_0-dy, z_0-dz), (x_0, y_0, z_0), and (x_0+dx, y_0+dy, z_0+dz) according to the concept of the virtual sound source. Here, dx, dy and dz can be calculated from a vector between a listener and the location of the sound source and the angle between the direction vectors of the sound source, the vector and the angle which are defined in an angle field and a direction field.

Fig. 2 describes a spatial sound source by using a plurality of point sound sources. Audio BIFS appears it can support the description of a particular scene. However, this method requires too much unnecessary sound object definition. This is because many objects should be defined to model one single object.

When it is told that the genuine object of hybrid description of Moving Picture Experts Group 4 (MPEG-4)is more object-oriented representations, it is desirable to combine the point sound sources, which are used for model one spatial sound source, and reproduce one single object.

In accordance with the present invention, a new field is added to a "DirectiveSound" node of the Audio BIFS to describe the shape and size attributes of a sound source. Fig. 3 shows an example where spatiality extension information is added to a "DirectiveSound" node of Audio BIFS in accordance with the present invention.

Referring to Fig. 3, a new rendering design corresponding to a value of a "SourceDimensions" field is applied to the "DirectiveSound" node. The "SourceDimensions" field also includes shape information of the sound source. If the value of the "SourceDimensions" field is "0,0,0", the sound source becomes one point, no additional technology for extending the sound source is applied to the "DirectiveSound" node. If the value of the "SourceDimensions" field is a value other than "0,0,0", the dimension of the sound source is extended virtually.

The location and direction of the sound source are defined in a location field and a direction field, respectively, in the "DirectiveSound" node. The dimension of the sound source is extended in vertical to a vector defined in the direction field based on the value of the "SourceDimensions" field.

The "location" field defines the geometrical center of the extended sound source, whereas the "SourceDimensions" field defines the three-dimensional size of the sound source. In short, the size of the sound source extended spatially is determined according to the values of Δx, Δy and Δz.

Fig. 4 is a diagram illustrating how a sound source is extended in accordance with the present invention. As illustrated in the drawing, the value of the "SourceDimensions" field is $(0, \Delta y, \Delta z)$, Δy and Δz being not zero ($\Delta y \neq 0$, $\Delta z \neq 0$). This indicates a surface sound source having an area of $\Delta y \times \Delta z$.

The illustrated sound source is extended in a direction vertical to a vector defined in the "direction" field based on the values of the "SourceDimensions" field, i.e., $(0, \Delta y, \Delta z)$, and thereby forming a surface sound source. As shown in the above, when the dimension and location of a sound source is defined, the point sound sources are located on the surfaces of the extended sound source. In the present invention, the locations of the point sound sources are calculated to be distributed on the surfaces of the extended sound source uniformly.

Figs. 5A to 5C are diagrams depicting the distributions of point sound sources based on the shapes of various sound sources in accordance with the present invention. The dimension and distance of a sound source are free variables. So, the size of the sound source that can be recognized by a user can be formed freely.

For example, multi-track audio signals that are

recorded by using an array of microphones can be expressed by extending point sound sources linearly as shown in Fig. 5A. In this case, the value of the "SourceDimensions" field is (0, 0, Δz).

Also, different sound signals can be expressed as an extension of a point sound source to generate a spread sound source. Figs. 5B and 5C show a surface sound source expressed through the spread of the point sound source and a spatial sound source having a volume. In case of Fig. 5B, the value of the "SourceDimensions" field is (0, Δy, Δz) and, in case of Fig. 5C, the value of the "SourceDimensions" field is (Δx, Δy, Δz).

As the dimension of a spatial sound source is defined as described in the above, the number of the point sound sources (i.e., the number of input audio channels) determines the density of the point sound sources in the extended sound source.

If an "AudioSource" node is defined in a "source" field, the value of a "numChan" field may indicate the number of used point sound sources. The directivity defined in "angle," "directivity" and "frequency" fields of the "DirectiveSound" node can be applied to all point sound sources included in the extended sound source uniformly.

The apparatus and method of the present invention can produce more effective three-dimensional sounds by extending the spatiality of sound sources of contents.

While the present invention has been described with respect to certain preferred embodiments, it will be apparent to those skilled in the art that various changes and modifications may be made without departing from the scope of the invention as defined in the following claims.